

Translation of Amended Claims

Filed: February 25, 2010

1. A method for optimizing a nucleotide sequence for the expression of a protein on the basis of the amino acid sequence of the protein, which comprises the following steps carried out on a computer:

generating, by a test sequence generating means, a first test sequence of  $n$  codons which correspond to  $n$  consecutive amino acids in the protein sequence, where  $n$  is a natural number and is less than or equal to  $N$ , the number of amino acids in the protein sequence;

specifying, by means for a determining optimization positions in a test sequence,  $m$  optimization positions in the test sequence which correspond to the position of  $m$  codons at which the occupation by a codon, relative to the test sequence, is to be optimized in relation to the test sequence, where  $m \leq n$  and  $m < N$ ;

generating, by a test sequence generating means, one or more further test sequences from the first test sequence by replacing at one or more of the  $m$  optimization positions a codon of the first test sequence by another codon which expresses the same amino acid at one or more of the  $m$  optimization positions;

assessing, by assessment means, each of the test

sequences with a quality function and ascertaining the test sequence which is optimal in relation to the quality function;

determining, by result codon determining means, p  
codons of the optimal test sequence as result codons,

which are located at one of the m  
optimization positions,

and which form the codons of the optimized nucleotide sequence at the positions which corresponds to the position of said p codons in the test sequence

where p is a natural number and  $p \leq m$ ;

iterating, by using iteration means, the  
preceding steps, where in each iteration step:

the step of generating a first test sequence  
comprises that for those amino acids in the amino acid  
sequence for which a result codon has been determined in a  
previous iteration the result codon determined previously  
is selected, and

the step of specifying m optimization  
positions comprises selecting the optimization positions in  
such a way that they do not coincide with the positions of  
amino acids in said amino acid sequence of a protein for  
which a result codon has been determined in a previous  
iteration.

2. The method as claimed in claim 1, characterized in that in one or more iteration steps the m optimization positions of the test sequences directly follow one or more

result codons which have been specified as part of the optimized nucleotide sequence.

3. The method as claimed in claim 1 or 2, characterized in that in one or more iteration steps the p codons which are specified as result codons of the optimized nucleotide sequence are p consecutive codons.

4. The method as claimed in any of claims 1 to 3, characterized in that in one iteration step test sequences with all possible codon occupations for the m optimization positions are generated from the first test sequence, and the optimal test sequence is ascertained from these test sequences.

5. The method as claimed in any of claims 1 to 4, characterized by:

assessment of each test sequence with a quality function,

ascertaining of an extreme value within the values of the quality function for all partial sequences generated in an iteration step,

specification of p codons of the test sequence which corresponds to the extremal value of the weight function as result codons at the appropriate positions, where p is a natural number and  $p \leq m$ .

6. The method as claimed in claim 5, characterized in that the quality function takes account of one or more of the following criteria:

codon usage for a predefined organism, GC content, repetitive sequences, secondary structures, inverse complementary sequence repeats and sequence motifs.

7. The method as claimed in claim 6, characterized in that the quality function is a function of various single terms which in each case assess one criterion from the following list of criteria:

codon usage for a predefined organism, GC content, sequence motifs, repetitive sequences, secondary structures, inverse complementary sequence repeats.

8. The method as claimed in any of claims 1 to 6, characterized in that the quality function takes account of one or more of the following criteria:

exclusion of inverse complementary sequence identities of more than 20 nucleotides to the transcriptome of a predefined organism,

exclusion of homology regions of more than 100 base pairs to a predefined DNA sequence,

exclusion of homology regions with more than 90% similarity of the nucleotide sequence to a predefined DNA

sequence.

9. The method as claimed in any of claims 1 to 8, characterized by the step of synthesizing the optimized nucleotide sequence.

10. The method as claimed in claim 9, characterized in that the step of synthesizing the optimized nucleotide sequence takes place in a device for automatic synthesis of nucleotide sequences which is controlled by the computer which optimizes the nucleotide sequence.

11. A device for optimizing a nucleotide sequence for the expression of a protein on the basis of the amino acid sequence of the protein, which has a computer means which comprises:

test sequence generating means for generating a first test sequence of  $n$  codons which correspond to  $n$  consecutive amino acids in the protein sequence, where  $n$  is a natural number and is less than or equal to  $N$ , the number of amino acids in the protein sequence;

means for a determining optimization positions in a test sequence for specifying  $m$  optimization positions in the test sequence which correspond to the position of  $m$  codons at which the occupation by a codon, relative to the test sequence, is to be optimized in relation to the test sequence, where  $m \leq n$  and  $m < N$ ;

test sequence generating means for generating one or more further test sequences from the first test sequence by replacing at one or more of the m optimization positions a codon of the first test sequence by another codon which expresses the same amino acid at one or more of the m optimization positions;

assessment means for assessing each of the test sequences with a quality function and ascertaining the test sequence which is optimal in relation to the quality function;

result codon determining means for determining p codons of the optimal test sequence as result codons, which are located at one of the m optimization positions,

and which form the codons of the optimized nucleotide sequence at the positions which corresponds to the position of said p codons in the test sequence where p is a natural number and  $p \leq m$ ;

iteration means for iterating the preceding steps, where in each iteration step:

the step of generating a first test sequence comprises that for those amino acids in the amino acid sequence for which a result codon has been determined in a previous iteration the result codon determined previously is selected, and

the step of specifying m optimization positions comprises selecting the optimization positions in

such a way that they do not coincide with the positions of amino acids in said amino acid sequence of a protein for which a result codon has been determined in a previous iteration.

12. The device as claimed in either of claims 11 characterized by a device for automatic synthesis of nucleotide sequences which is controlled by the computer in such a way that it synthesizes the optimized nucleotide sequence.

13. A computer program which comprises program code which can be executed by a computer and which, when it is executed on a computer, causes the computer to carry out a method as claimed in any of claims 1 to 8.

14. The computer program as claimed in claim 13, where the program code can, when it is executed on a computer, cause a device for the automatic synthesis of nucleotide sequences to prepare the optimized nucleotide sequence.

15. A computer-readable data medium on which a program as claimed in either of claims 13 or 14 is stored in computer-readable form.